

画像とテキストの利用による同一ニュースの言語横断検索

小川 晃[†] 高橋 友和^{†,††} 井手 一郎^{†,†††} 村瀬 洋[†]

[†] 名古屋大学大学院 情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

^{††} 日本学術振興会

^{†††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-1

E-mail: [†] {aogawa,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ^{†††} ide@nii.ac.jp

あらまし 近年、ニュース映像のアーカイビングが重要となり、大量に蓄積されたニュース映像の整理や閲覧の効率化のためにトピックの検索・追跡技術が必要とされている。また、最近では、あるニュースの様々な観点からの内容理解などのため、国内だけでなく海外ニュース映像を閲覧する機会が増加している。そこで、我々は放送国の異なるニュース映像から同一ニューストピックを検索する技術に注目する。しかし、従来のテキストのみを用いた手法では機械翻訳の性能や放送国間の視点・文化的背景の違いにより十分な検索は困難である。本研究では、テキスト情報に加え、映像中の画像情報を用いてニューストピック間の類似性を決定することで、言語の違いにも頑強な同一ニュース検索を行なう手法を提案する。画像情報としては、映像間における同一素材映像の存在を用いる。実験の結果、テキスト情報を補う情報としての画像情報の有用性と画像・テキストの併用による検索精度の向上を確認した。

キーワード 言語横断情報検索, 映像検索, 同一ニュース, 同一映像区間

Cross-lingual retrieval of identical news by using image and text information

Akira OGAWA[†], Tomokazu TAKAHASHI^{†,††}, Ichiro IDE^{†,†††}, and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

^{††} Japan Society for the Promotion of Science

^{†††} National Institute of Informatics 2-1-1 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: [†] {aogawa,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ^{†††} ide@nii.ac.jp

Abstract Recently, the importance of archiving news videos has increased. Therefore, for reusing large quantities of accumulated news videos efficiently and effectively, technology for news topic retrieval and tracking is necessary. Moreover, in order to understand a certain topic from various viewpoints, the opportunity for viewing news programs from all over the world has increased. Therefore, we focus on identical event detection in various news programs from different countries. However, cross-lingual retrieval using text information is difficult due to the performance of machine translation and moreover, different viewpoints and cultures. In this paper, we propose a cross-lingual retrieval method for detecting identical news topics that exploits image information together with text information. We utilize the existence of near-duplicate video segments as a strong information for the purpose. In an experiment, we verified the effectiveness of making use of the existence of near-duplicate video segments and the possibility of improving retrieval performance by using them.

Key words Cross-lingual information retrieval, video retrieval, identical news, near-duplicate video segments

1. はじめに

1.1 背景と目的

近年、記憶装置や通信技術の発達により、我々は数 1,000 時

間規模の映像データベースを容易に構築できるようになりつつある。放送映像にはスポーツ、ドラマ、バラエティ、アニメーションなど様々な種類があるが、なかでもニュース映像は人間社会の出来事を克明に記録した映像データであり、資料的価値

が高いため、映像のアーカイビングが重要視されている。しかし、複数の放送局で毎日放送されるという特性があり、その蓄積量は非常に膨大なものになる。それ故、それら大容量のニュース映像を効率的に整理・閲覧するため、話題（トピック）の検索・追跡といった技術が必要とされている。

また最近では、衛星放送や Web 上での映像配信などにより、国内の放送映像だけでなく海外の放送映像を視聴することも容易になりつつある。ニュース映像においては、様々な視点から情勢を理解するための資料として、世界各国のニュース映像の利用価値が高い。このようなニーズは、国際化が進むにつれて増加すると考えられるので、国内のニュース映像に加えて海外の映像も映像資源として利用する技術の研究が必要である。

以上のことから、世界各国のニュース映像を対象とした話題の検索・追跡技術の重要性は高く、それらを実現するために、異なる言語・放送国間で関連した話題を扱ったニュースを対応付ける技術が求められている。そこで、本研究では、異なる放送国の様々なニュース映像から同一ニューストピックを検索する技術に注目する。

現在、一般にニュース映像を対象とした検索は、新聞記事検索などのテキスト検索技術を拡張し、音声の書き下しや TV 画面上に現れるテロップを用いたテキストのみによる手法が用いられている。このような手法は、同一言語・放送国のニュース映像のみを対象とした検索においては比較的容易に実現することができ、検索精度も実用的な高さである。しかし、前述したとおり映像資源としては国内のものだけでなく海外の放送映像も存在するため、それら映像資源を含めて網羅的に利用することが望まれる。つまり、検索対象となるニュース映像が、異なる複数の言語・放送局から構成されるようになる。このような映像アーカイブに対して従来のテキストのみによる手法で検索を行なう場合、機械翻訳の性能の問題や国や放送局による視点・文化的背景の違いにより、一般的に同一言語を対象としたときより検索性能が低下する。また、全てのニュース映像に必ずしも文字放送字幕（Closed Caption; CC）テキストのような音声を書き下したテキストが付随しているわけではないので、音声認識によるテキスト情報を用いなければならない場合もある。しかし、現在の音声認識技術の性能では、周囲の雑音などの影響で正しく音声を書き下すことができないこともあり、十分なテキスト情報を得ることができない。

そこで本研究では、テキスト情報に加えてニュース映像中の画像情報を利用する手法を提案し、異なる言語・放送国間での同一ニューストピック検索の精度向上を目指す。一般にニュース映像では、世界規模、もしくは近隣国の話題を報道する場合、その国の放送局などから素材映像の提供をうけることが多く、希少な素材映像に関しては放送国・局を問わずに配信されて放送される性質がある。また、長期間にわたってある特定の話題を報道する際に、同一素材映像を象徴的に使いまわす傾向がある。これらを考慮すると、異なるニュース映像間に同一素材映像が含まれている場合、それらニュースの間には何らかの意味的関連性があるため、同一の話題を扱っている可能性があると考えられる。

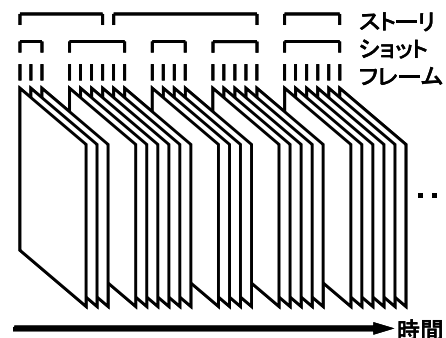


図 1 映像の構成

以上のことより我々は、画像情報としてニュース映像間における同一素材映像の存在を利用し、画像とテキストを併用した同一ニューストピックの検出手法を提案する。そして、その結果得られた検出精度をテキスト情報のみによる同一ニューストピック検出の精度と比較することで、画像情報がテキスト情報を補う情報として有用であり、検出精度の向上に利用できるかを検証する。

以降、2. 章で本研究に関連する研究について述べ、3. 章で提案手法についての詳細を述べる。4. 章で提案手法の効果を実験によって示し、5. 章で実験家かを用いた閲覧インタフェースについて述べる。6. 章でまとめる。

1.2 用語の整理

まず、関連する用語について整理をする。本研究において、ニュース映像とは動画と音声の集合体であり、存在する場合には CC テキストもこれに含む。一方、素材映像とは TV 局により編集が行なわれる前の映像のことである。また、映像の画像的構成は

- フレーム：動画の最小構成単位である静止画像
- ショット：画像的に連続するフレーム群

とし、内容的構成は

- ストーリー：ニュース映像中で 1 つのイベントを扱う単位
- イベント：ある日ある場所で起こった出来事

とする（図 1）。

2. 関連研究

ニュース映像を対象とした検索・追跡に関連した研究には以下のようなものが挙げられる。

角田ら [2] は、ニュース映像中のストーリーに対応する新聞記事をテロップと記事のテキスト照合により特定する手法を提案した。この研究は、ニュース映像の画面に表れるテロップと新聞記事のテキストを用い、一致文字列の長さや位置、頻度などを用いてニュースの報道のそれぞれに対応する新聞記事を 1 つ選択する手法である。同様の研究で、渡辺ら [1] は、TV ニュースと新聞が相補的な特徴を持っていることに注目し、対応関係あるいは続報関係にある TV ニュースの映像と新聞記事を取り出す手法を提案した。この手法ではテロップから名詞を抽出し、それをキーとしてニュース映像中のストーリーと新聞記事の類似度を求め、対応関係または続報関係にあるかどうかを判断して

いる。そして、その結果をデータベース化することで、ニュース映像と新聞記事を対象とした相互閲覧および検索システムを作成している。

ニュース映像の追跡に関する研究では、井手ら [3] が、大規模ニュース映像アーカイブ中のトピックのスレッド構造を抽出する手法と、その構造に基づく閲覧インタフェースを提案した。この研究では、CC テキストを用い、名詞列をキーとして、各ニュース映像をストーリー分割し、全てのストーリー間の類似度を評価することでトピックの追跡を行ない、スレッド構築を行っている。そして、得られたスレッド構造に基づき、映像閲覧インタフェース“Topic Browser”を実装している。また、追跡技術においては、Topic Detection Tracking (TDT) と呼ばれるオンラインニュースやニュース放送といったデータストリームからトピック構造を自動的に得るための技術の確立を目的とした技術研究開発プログラムがあり、そこで様々な研究がなされている。

これらの研究は、テキスト情報のみを用いた手法となっているが、高度な検索・追跡を実現している。しかし、それはあくまでも同一言語のニュース映像、新聞記事を対象としている場合であり、複数言語間を対象とした言語横断型の検索・追跡に適用する場合、機械翻訳の性能に関する問題や放送国間の異なる視点・文化的背景などにより、精度が低下することが知られている。TREC (Text REtrieval Conference) や NT-CIR(NII/NACSIS Test Collection for Information Retrieval) といったワークショップでは言語横断検索に対する研究も多くなされているが、そのほとんどはテキスト情報のみを用いたものである。

そこで本研究では、テキスト情報に加え、ニュース映像中の画像情報として同一素材映像の存在を用いることで、同一ニューストピック検索の精度向上を目指す。特定の区間映像に対し、同一映像を長時間映像群から検索する手法としては時系列アクティブ探索法 [4] をはじめ、いくつかの手法が存在する。しかし、ニュース映像間から任意の同一映像区間を検出する技術はいまだ確立されていない。本研究では、ニュース映像間における同一映像区間の高速検出手法として、野田らが提案した主成分分析を用いた特徴次元圧縮による同一映像区間の高速検出手法 [5] を用いる。この手法は、空間方向と時間方向の次元圧縮により、低次元の特徴量で同一映像区間を絞り込むことで検出精度を保ったまま高速に同一映像区間を検出するものである。

3. 提案手法

3.1 手法の概要

本研究では映像間における同一素材映像の共起性に注目し、従来のテキスト情報に加えて画像情報を用いる同一ニューストピックの検出手法を提案する。この手法は、画像・テキスト情報それぞれにおける映像間の類似性を組み合わせることで同一ニューストピックを検出する。検出処理の流れは図2のとおりである。まず、準備処理として各ニュース映像をストーリーに分割する。次に、各ストーリー間で同一素材映像の共起に基づく画

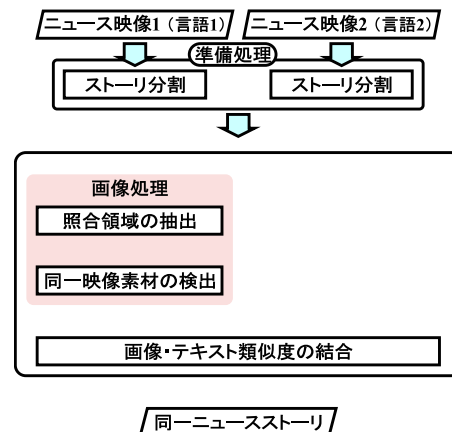


図2 提案手法による同一ニューストピック検出の処理手順

像の類似度を、キーワードの出現頻度を用いてテキストの類似度を求める。最後に、それら2つを組み合わせた類似度を用いて同一ニューストピックを検出する。以下、それぞれの処理について詳しく説明する。

3.2 準備処理

一般にニュース映像は、複数のストーリーにより構成されたストーリーの集合体である。そこで本研究では、ニュース映像をストーリーごとに分割し、それを同一ニューストピックを検出する単位とする。ニュース映像におけるストーリー分割としては、CCテキストを用いたテキスト処理による手法やキャスタショットの出現を用いた画像処理による手法など、従来より様々な研究が行われている。そこで、本研究ではそれら従来手法を用いてストーリー分割を行なうことを前提にする。以下の実験において、日本語のニュース映像に関しては、CCテキストにおける文間のキーワードベクトルの類似性に基づく手法 [3] により分割された結果を用いる。一方、海外のニュース映像として用いる、米国 NIST 主催の TrecVid2005 [11] の映像データの一部に対して、文献 [9] などによりストーリー分割された結果が提供されている参加者共有データを用い、残りは一部人手による結果を用いた。

3.3 画像処理

3.3.1 照合領域の抽出

一般にニュース映像は、番組ごと独自のテロップを挿入するなどの加工がされているため、同じ映像素材を基にしている場合でも部分的に異なる映像となり、類似度が低下して検出ができない恐れがある。この問題を回避するための単純な方法として検出時の類似度の閾値を低くすることが考えられるが、類似度の低下に比例して誤検出が増加すると考えられる。大規模なニュース映像アーカイブに適用する場合、誤検出が多くなることは望ましくないため、本手法では比較的テロップなどが挿入されることが少ない各フレームの中心領域を切り出し、その領域のみを照合に用いることにする(図3)。切り出し位置は、目視で大量のニュース映像から決定した。

また、ニュース映像間の照合における問題は、映像の加工による違いだけではなく、録画状況や各放送局の規定による映像の明るさの違いに関するものも考えられる。本手法では特徴量



図 3 照合領域の切り出しと明るさの正規化

として各画素の RGB 値を使用するため、この変化により検出漏れや誤検出が生じる可能性が高い。そこで、ニュース映像中の各照合領域の輝度値に対し、ヒストグラム平坦化 [6] を施すことで明るさの正規化を行ない、この問題に対処する (図 3)。本手法では、RGB 値を YUV 値に変換し、輝度値である Y に対して平坦化を行い、再び RGB 値に戻すことで正規化を行なう。以降、RGB 値としてはこの正規化により得られた各画素の RGB 値を用いて処理を行なう。

3.3.2 同一素材映像の検出

本研究では、ニュース映像中から同一素材映像を検出する手法として、特徴次元圧縮による長時間映像中における同一区間映像の高速検出手法 [5] を用いる。この手法は、フレーム画像中の各画素の RGB 値を特徴ベクトルとし、空間方向と時間方向の次元圧縮により低次元化された特徴ベクトルを用いて繰り返し照合することで、検出精度を保持したまま同一映像区間の候補を高速に検出する。そして、候補として得られた映像区間に対し、次元圧縮する前の高次元の特徴ベクトルによる詳細照合により、同一映像区間を正確に検出する。

3.4 テキスト処理

この節では、テキスト情報によるニュースストーリー間の類似度を求める処理について説明する。テキスト情報として CC テキストを用いて処理を行なう。

まず、和文以外の CC テキストに関しては、それぞれの言語に適した既存の翻訳ソフトウェアや Web 上の翻訳サービスを用いて機械翻訳を行ない、和文のテキストに変換して用いることを前提とする。以下の本実験では、外国の映像は米国のニュース映像のみを用いたので、翻訳ソフトウェア「The 翻訳プロフェッショナル v10」[東芝]を用いた。次に、日本語形態素解析システム JUMAN [7] を用いて CC テキストを形態素解析し、品詞判別を行なう。そして、得られた形態素の品詞情報を用いて keyword となる語句を抽出する。本実験においては、以下に示す品詞を持つ語句を keyword とする。

- 名詞:「人称名詞」や「形式名詞」は対象外とし、前後の「接頭辞」、「接尾辞」は結合、名詞が連続する場合はそれらを結合した名詞列も 1 つの名詞として扱う

- 未定義語:「地名」や「人名」を表す語のみ

以上の処理により得られた keyword とその出現頻度を用い

表 1 実験に用いたニュース映像の詳細

国名	放送局	番組名	放送時間
日本	NHK	News7	19:00 ~ 19:30
米国	CNN	NewsNight Aaron Brown	22:00 ~ 23:00

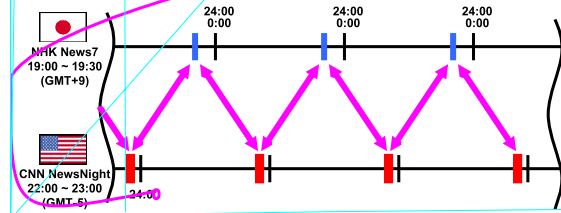


図 4 放送時間と映像グループの関係

て各ニュースストーリー間の類似性を求める。類似度としては、ストーリー間の keyword ベクトルのコサイン距離を用いる。

4. 実験

言語横断型の同一ニューストピック検索における画像情報の有用性と、画像とテキストの併用による同一ニューストピック検索の精度向上を確認するために、提案手法を用いて異なる言語・放送国のニュース映像間から同一ニューストピックを検出する実験を行なった。

4.1 実験対象

検出対象として 2004 年 11 月に実際に放送された日米の 2ヶ国のニュース映像を使用した。それぞれの放送局と番組名、放送時間を以下に示す。

なお、米国のニュース映像は TrecVid2005 [11] 提供のものを利用し、CC テキストは CNN が Web 上で公開しているもの [12] を用いた。また、日米間の時差やニュース伝達の時間を考慮し、日本で放送されたニュース映像 1 本に対し、その前後 24 時間以内に米国で放送されたニュース映像 2 本に照合対象を限定した (図 4 に示す映像グループ)。本実験では 18 の映像グループを用いて検出を行なった。

4.2 同一素材映像の検出精度

本研究では、画像情報として同一素材映像の存在を用いているため、画像情報による同一ニューストピック検出の精度を評価するうえで、同一素材映像の検出精度が重要となる。そこで本節では、前節で述べた実験対象における同一素材映像の検出精度について述べる。

3.3 節で述べた手法により検出を行なった結果、適合率は 93% (79/85)、再現率は 81% (79/98) となった。適合率に関しては、本実験では誤検出の削減のため、検出された同一素材映像に対してさらに色コリログラム [8] による絞込み照合を行なったこともあり、90% 以上を示し、再現率においても 80% 以上を示したので、検出精度としては十分であると考えられる。なお、誤検出について分析したところ、低解像度化や輝度のヒストグラム平坦化の影響により、照合領域がそれぞれ海と空を表した映像のように、映像素材は異なるが画面的 (特に色特徴的) には違いがあまりない映像同士を誤対応してしまうものであった。また、検出漏れについても分析したところ、映像間の

図 6 映像閲覧インタフェース

で、再びそれらの中から閲覧したいトピックを選択することができる。

この閲覧インターフェースはまだ開発途中のものであり、実用性という点で表示される情報や機能が不足していると考えられるので、今後改善が必要である。

6. む す び

本研究では、映像の検索技術の 1 つとして、異なる言語・放送国間における同一ニューストピックの検出を行なった。従来研究で行なわれているようなテキスト情報のみによる検出ではなく、画像情報としてニュース映像間の同一素材映像の存在を利用し、テキスト情報と併用することで、使用言語の違いや文化的背景の違いなどの問題により困難であった言語横断型の検出を同一言語・放送国を対象としたときと同等の精度で行なえる手法を提案した。本実験では、実際に放送された日本とアメリカのニュース映像に対し、提案手法による同一ニューストピック検出を行ない、その結果、テキスト情報では検出することができなかった同一ニューストピックを画像情報を用いることで検出することができることを確認した。また、画像情報とテキスト情報を併用して検出を行なうことで、言語横断型同一ニューストピック検出の精度向上に成功した。

本研究で得られる同一ニューストピックは映像の検索以外にも、閲覧や整理、要約にも応用できる。例えば、あるニュースに注目しているときに、同じトピックを扱っている他のニュースを検索・閲覧することや、あるトピックに関する世界中のニュース映像を集めることで、より深い内容の要約や映像の補完などができると考えられる。

今後の課題としては、本実験では画像とテキストを併用する際に両情報を同等にして扱ったが、話題の種類による映像の提供具合や報道内容における視点・文化的背景の影響を考慮した重み付け、より効果的な画像・テキストにおける類似度の定義や組み合わせ方の検討が必要である。

また、画像情報については同一素材映像だけでなく、会見などでの撮影カメラ位置の違いや映像の拡大縮小に口バストで、かつ高速に検索できる特徴の検討・導入が挙げられる。類似映

像の検出においては、すでにキーフレーム画像の SIFT 特徴を用いた手法 [10] が提案されているが、この手法は静止画による照合であり、照合における計算時間も膨大であるため、実用性を考えるとそのまま用いることは難しい。

そして最終的には、様々な国のニュース映像を含む、より大規模な映像アーカイブに適用することが望まれる。

謝 辞

実験のデータとして使用したニュース映像を提供して頂いた国立情報学研究所、米国 National Institute of Standards and Technologies による TREC Video 2005 ワークショップに感謝する。日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する。本研究の一部は文部科学省科学研究費補助金による。本研究では、画像処理に MIST ライブラリ (<http://mist.suenaga.m.is.nagoya-u.ac.jp/>) を使用した。

文 献

- [1] 渡辺靖彦, 岡田至弘, 金地健吾, 阪元慶隆: “TV ニュースと新聞記事を対象にしたマルチメディアデータベースシステム”, 信学技報, PRMU97-257, Mar. 1998.
- [2] 新谷研, 角田達彦, 大石巧, 長尾真: “キャプションと記事テキストの文字列照合による報道番組と新聞記事との対応づけの自動化”, 情報処理学会論文誌, vol.38, no.6, pp.1149–1167, June 1997
- [3] Ichiro Ide, Hiroshi Mo, Norio Katayama, Shin'ichi Satoh: “Exploiting topic thread structures in a news video archive for the semi-automatic generation of video summaries”, Proc. 2006 IEEE Intl. Conf. on Multimedia and Expo (ICME2006), pp.1473–1476, July 2006.
- [4] 柏野邦夫, 黒住隆行, 村瀬洋: “ヒストグラム特徴を用いた音や映像の高速 AND/OR 探索”, 信学論, vol.J83-D-II, no.12, pp.2735–2744, Dec. 2000.
- [5] Ichiro Ide, Kazuhiro Noda, Tomokazu Takahashi, Hiroshi Murase: “Genre-adaptive near-duplicate video segment detection”, Proc. 2007 IEEE Intl. Conf. on Multimedia and Expo (ICME2007), pp.484–487, July 2007.
- [6] Graham Finlayson, Steven Hordley, Gerald Schaefer, Gui Yun Tian: “Illuminant and device invariant colour using histogram equalisation”, Pattern Recognition, vol. 38, issue 2, pp.179–190, Feb. 2005.
- [7] 黒橋禎夫, 河原大輔: “日本語形態素解析システム JUMAN version5.1”, 東京大学大学院情報理工学系研究科, <http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html> より入手, Sept. 2005.
- [8] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, R. Zabih: “Image indexing using color correlograms”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition '97, pp.762–768, June 1997.
- [9] Yun Zhai, Alper Yilmaz, Mubarak Shah: “Story segmentation in news videos using visual and text cues”, Proc. Fourth Int. Conf., CIVR 2005, pp. 92–102, July 2005.
- [10] Chong-Wah Ngo, Wan-Lei Zhao, Yu-Gang Jiang: “Fast tracking of near-duplicate keyframes in broadcast domain with transitivity propagation”, Proc. the 14th annual ACM international Conf. on Multimedia, pp. 845–854, Oct. 2006.
- [11] National Int., “TrecVid2005 Evaluation”, <http://www-nlpir.nist.gov/projects/tv2005/tv2005.html>
- [12] CNN.com TRANSCRIPTS, <http://transcripts.cnn.com/TRANSCRIPTS>